

REGION-SPECIFIC ROLE OF WATER IN COLLAGEN UNWINDING AND
ASSEMBLY

A Thesis

by

KRISHNAKUMAR MAYURAM RAVIKUMAR

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2008

Major Subject: Biomedical Engineering

REGION-SPECIFIC ROLE OF WATER IN COLLAGEN UNWINDING AND
ASSEMBLY

A Thesis

by

KRISHNAKUMAR MAYURAM RAVIKUMAR

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE

Approved by:

Chair of Committee,	Wonmuk Hwang
Committee Members,	Jay D. Humphrey
	Alvin T. Yeh
	Mariappan Muthuchamy
Head of Department,	Gerard L. Côté

August 2008

Major Subject: Biomedical Engineering

ABSTRACT

Region-specific Role of Water in Collagen Unwinding and Assembly. (August 2008)

Krishnakumar Mayuram Ravikumar, B.E., University of Madras

Chair of Advisory Committee: Dr. Wonmuk Hwang

Conformational stability of the collagen triple helix affects its turnover and determines tissue homeostasis. Although it is known that the presence of imino acids (prolines or hydroxyprolines) confer stability to the molecule, little is known regarding the stability of the *imino-poor* region lacking imino acids, which plays a key role in collagen cleavage. In particular, there have been continuing debates about the role of water in collagen stability. We addressed these issues using molecular dynamics simulations on 30-residue long collagen triple helices, including a structure that has a biologically relevant 9-residue imino-poor region from type III collagen (Protein Data Bank ID: 1BKV). We characterized the conformational motion of the molecule that differs between imino-rich and imino-poor regions using a torsional map approach. At temperatures of 300 K and above, unwinding initiates at a common cleavage site, the glycine-isoleucine bond in the imino-poor region. This provides a linkage between previous observations that unwinding of the imino-poor region is a requirement for collagenase cleavage, and that isolated collagen molecules are unstable at body temperature. Unwinding of the imino-poor region is controlled by dynamic water bridges between backbone atoms with average lifetimes on the order of a few picoseconds, as the degree of unwinding strongly correlated with the loss of water bridges, and unwinding could be either prevented or enhanced, respectively by enforcing or forbidding water bridge formation. While individual water bridges were short-lived in the imino-poor region, the hydration shell surrounding the entire molecule was stable even at 330 K. The diameter of the hydrated collagen including

the first hydration shell was about 14 Å, in good agreement with the experimentally measured inter-collagen distances. These results elucidate the general role of water in collagen turnover: water not only affects collagen cleavage by controlling its torsional motion, but it also forms a larger-scale lubrication layer mediating collagen self-assembly.

ACKNOWLEDGMENTS

We would like to thank Jan K. Rainey for providing the THeBuScr collagen building script ([http:// structbio.biochem.dal.ca/jrainey/THeBuScr.html](http://structbio.biochem.dal.ca/jrainey/THeBuScr.html)).

TABLE OF CONTENTS

CHAPTER		Page
I	INTRODUCTION	1
II	METHODS	6
	A. Simulation of 1BKV	6
	B. GPP and GPO peptides	8
	C. Characterization of unwinding	9
	D. Water bridging and pertinent simulations	10
III	RESULTS	12
	A. Imino-poor region is more prone to unwinding	12
	B. Dynamics of surface water bridges	13
	C. Loss of water bridges causes unwinding	18
	D. Dual role of water as stabilizer and as thermal agitator	22
	E. Organization of hydration layers depends on the imino- acid content	23
	F. The first hydration shell defines the boundary for col- lagen self-assembly	29
IV	DISCUSSION AND CONCLUSION	32
	A. Discussion	32
	B. Conclusion	35
	REFERENCES	37
	APPENDIX A	44
	VITA	47

LIST OF TABLES

TABLE		Page
I	Summary of simulations performed. In simulations with no charge on select backbone H and/or O atoms, water bridge formation is prevented. At 330 K, k is the spring constant of the harmonic restraint used to enforce water bridges ($1 \text{ kcal}/(\text{mol } \text{\AA}^2) = 695 \text{ pN/nm}$). The reflecting boundary prevents the peptide from diffusing out of the simulation box in longer 5 ns simulations.	7

LIST OF FIGURES

FIGURE		Page
1	Backbone hydrogen bonds and water bridges in the imino-poor region of 1BKV. Chain 1 repeats after chain 3 to show bonds between the two. Side chains are not shown and Gly is shaded gray. Dotted: direct backbone hydrogen bonds. Dashed with the letter ‘W’: surface water bridges, named from B1 to B9. Note that an imino acid does not have the backbone N-H group, thus cannot form a water bridge if it occupies the position X. Each triad consists of three residues to the right of the triad number (<i>see</i> , Fig. 2). Triads outside the range 5-14 are in imino-rich domains (not shown).	2
2	Characterization of unwinding using triads. Each triad is based on the triangle that connects neighboring C_α atoms (represented by spheres) forming the cross section of the molecule. A set of arrows attached to each triangle represent the local coordinate basis of the triad, whose torsional angle can be used to quantify helical twist and unwinding [17]. Conformations are shown: (A) At start and (B) after 5 ns of MD at 330 K. Triads 5, 10, and 15 are numbered. The unwound region is marked in (B).	10
3	Temperature-dependent unwinding. The local slope of each curve is directly proportional to the helicity at the corresponding triad. Each point corresponds to the helical angle difference from the first triad averaged over successive 0.4-ns or 1-ns intervals (\pm standard deviation). Triads 5-8 contain Gly-Ile bonds, where unwinding initiates.	14
4	Torsional map of imino-rich peptides at 300 K (<i>see</i> , Fig. 3). (A) (GPP) ₁₀ and (B) (GPO) ₁₀ . Note that $\Delta\theta_z$ at triad 24 is larger than those in Fig. 3 , due to their tighter twist.	15

FIGURE

Page

5	Analysis of water bridging events. (A) Average lifetime and occupancy (occupancy time divided by the simulation time) of water bridges. Standard deviations in lifetimes are 3.03 ps (273 K), 1.74 ps (300 K), and 1.44 ps (330 K). Note that standard deviations are larger than averages, indicating that the distribution of lifetimes is asymmetric with a long tail at large lifetimes. (B) Average lifetimes and (C) occupancy of individual water bridging sites (<i>cf.</i> , Fig. 1).	16
6	Average occupancy of individual water bridges. The total simulation time at 273 K (2 ns), 300 K (5 ns), and 330 K (5 ns) was divided into five intervals as shown in the graph. Bridge numbers B1-B9 are from Fig. 1	19
7	Correlation between the number of water bridges and the helical angle difference across the imino-poor region (triads 5-15). The angle difference was normalized within the 20°-180° window and the number of water bridges was normalized within the 0-9 window. (A) 273 K, (B) 300 K, and (C) 330 K	20
8	Suppression of unwinding by enforcing water bridge formation. (A) Torsional map (<i>see</i> , Fig. 3). (B) Number of water bridges and the helical angle difference across the imino-poor region (triads 5-15; <i>see</i> , Fig. 7).	21
9	Preventing water bridge formation. A bridging site with amino hydrogen and carboxylic oxygen from two adjacent chains along with a water molecule is shown. Signs of point charges (+, −, or 0) are marked above respective atoms. (A) Normal case when a water bridge is formed. (B) The point charge on N-H of X in the imino-poor region is removed, preventing bridge formation. But water can form a hydrogen bond with C=O of Gly (Case 1). (C) In addition to (B), charge on C=O of Gly is removed (Case 2). . . .	23
10	Torsional map of 1BKV in the absence of water bridges. (<i>cf.</i> , Fig. 3). (A) Case 1: No charge on N-H of X. Unwinding was observed even at 273 K. (B) Case 2: No charge on N-H of X and C=O of Gly. Unwinding was not observed during the simulation. . .	24

FIGURE

Page

11	Hydration structure around the molecule. (A-C) Snapshot of water oxygens (at 970 ps; an arbitrarily chosen time) at (A) 273 K, (B) 300 K, and (C) 330 K. Water oxygens are closer to the backbone in the imino-poor region compared to imino-rich regions. (D-F) Average radial distribution of water oxygen atoms over the simulation period (273K: 2ns; 300,330 K: 5ns). The overall profile is preserved during the simulation, although the peak height decreases slightly as temperature increases (see also Fig. 14). (D) Imino-rich region, (E) imino-poor region, and (F) both regions together.	25
12	Position of water hydrogens around collagen at 970 ps at 273 K. Note that the first shell of water hydrogens ($\leq 2.00\text{\AA}$) are closer to the backbone than water oxygens in the imino-poor region [Fig. 11(A-C)]. This reflects the shorter length of the hydrogen bond between water hydrogen and backbone oxygen compared to that between water oxygen and backbone hydrogen, due to the stronger electrostatic interaction (a larger backbone point charge) in the former case.	26
13	A cartoon representing the orientation of water molecules (W1–W4; not to scale). Black circle: oxygen. White circle: hydrogen. Dashed lines (– –) separate important hydration regions including the two peaks and the void region in Fig. 11(D-F) . W1 and W3: Typical orientations of water bridges found only in the imino-poor region, where one O-H bond is approximately parallel to the surface of collagen. W2 and W4: Typical orientations radially perpendicular to the backbone, which are present in both imino-poor and imino-rich regions. Vertical arrows show approximate distances of hydration peaks from the backbone of 1BKV. Note that only atoms that lie within the first peak ($\leq 2.4\text{\AA}$) can form hydrogen bonds with the backbone.	28
14	Average radial distribution function of water oxygens at different times and temperatures. Column 1: imino-rich, column 2: imino-poor, column 3: both regions together.	31

FIGURE

Page

15	The three α -chains are shown in blue (chain-1), red (chain-2), and yellow (chain-3). Spheres are C_α atoms of Gly (green) and X/Y (orange). (A) Overview of 1BKV with the 24 triads. (B) Magnified view of the box in (A). To show the face of the triangle, the molecule is rotated about the vertical axis of the paper plane. Two neighboring triads are denoted $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ and $\{\hat{e}'_1, \hat{e}'_2, \hat{e}'_3\}$. (C) Triad numbers versus amino acid sequence. The horizontal shift of the three chains reflects their staggered structure in the triple helix. Triads are grouped into domains I, II, and III, as explained in the text. T1 and T2 are unconsidered regions to eliminate end effects. Triads 5 to 8, marked between grey dotted vertical lines, contain Gly-Ile bonds. (Note: The first proline (P) residue in chain-1 in (C) is shown only for clarity and is invisible in the X-ray crystal structure of 1BKV.)	46
----	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

CHAPTER I

INTRODUCTION

The collagen triple helix is a fundamental structural motif composed of three polypeptide II type α chains wound in a right-handed manner. The primary structure of each α chain consists of GXY triplets (G: Gly and X,Y: other amino acids). The triple helical structure has been studied for several decades, and features that stabilize the molecule without specifically involving water molecules are relatively well-established. Hydrogen bonds are formed between N-H of Gly and C=O of X in adjacent α chains along the length of the molecule [1], following the model II of Rich and Crick [2] (**Fig. 1**). The molecule is further stabilized by imino acids in X (Pro) or Y (Pro or Hyp: hydroxyproline), that provide a structural constraint giving a tightly wound tertiary structure [3]. However, the role of water in collagen stability is still under debate. The enhanced stability conferred by Hyp in the Y position was initially thought to be mediated by additional water bridges formed with the -OH group of Hyp, as observed in x-ray structures [4, 5, 6, 7].

But Engel *et al.*, observed that the triple helix formed by the (GPO) $_n$ peptide (O: one-letter code for Hyp) was more stable than that by the (GPP) $_n$ peptide in non-aqueous solutions, suggesting that higher stability conferred by Hyp is not mediated by specific water bridges [8]. Raines and co-workers proposed a different explanation [9, 10]. Rather than by water, they showed that the inductive effect of the -OH group in Hyp favors a more stable *trans* conformation. Later studies by Slatter *et al.*, however, suggested that the inductive effect cannot be solely responsible for collagen stability, but staggering of α chains and external bonds, likely mediated by water,

This thesis follows the style of *IEEE Transactions on Automatic Control*.

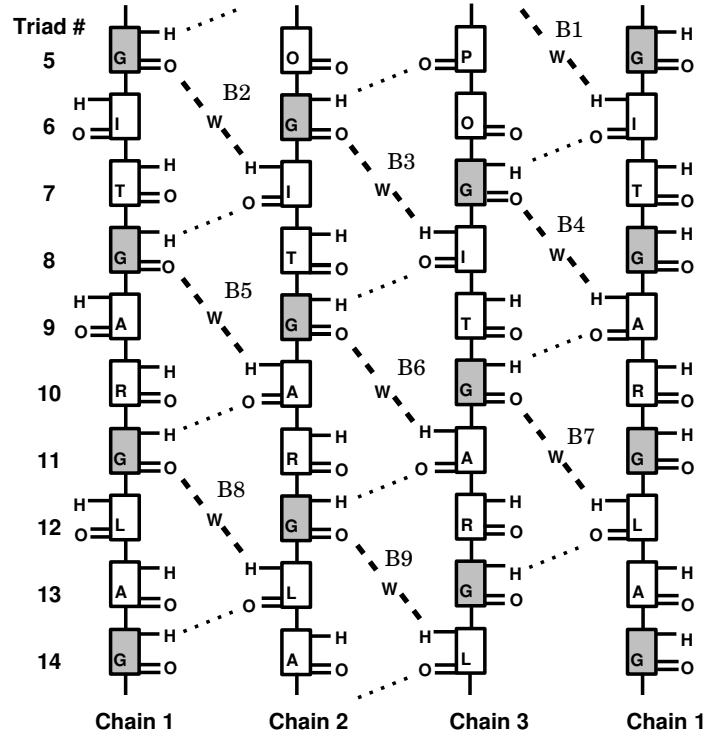


Fig. 1. Backbone hydrogen bonds and water bridges in the imino-poor region of 1BKV. Chain 1 repeats after chain 3 to show bonds between the two. Side chains are not shown and Gly is shaded gray. Dotted: direct backbone hydrogen bonds. Dashed with the letter 'W': surface water bridges, named from B1 to B9. Note that an imino acid does not have the backbone N-H group, thus cannot form a water bridge if it occupies the position X. Each triad consists of three residues to the right of the triad number (*see*, **Fig. 2**). Triads outside the range 5-14 are in imino-rich domains (not shown).

might be necessary [11].

In any case, more than 60% of GXY triplets in collagen do not have Hyp in the Y position [12], thus it is unclear to what extent Hyp contributes to the stability of the entire molecule. Furthermore, when the entire type I collagen instead of the GPP or GPO peptides was tested in alcoholic solutions, there was a non-monotonic dependence of its stability on the concentration of alcohol [13], suggesting that collagen stability is affected by hydration in a complex way.

Even less is clear about the stability of the imino-poor region (the region devoid of imino acids). Only in these regions, x-ray crystallography found bridges formed by a single water molecule between N-H of X and C=O of Gly in the adjacent α chain [3, 14, 15] (**Fig. 1**). These bridges likely form and break in solvated collagen molecules, and a static picture based on the crystal structure does not provide enough evidence for their stabilizing role. Stability of the imino-poor region is particularly important for the biological function of collagen. These regions are called *thermally labile domains* [16, 17], which are more flexible than imino-rich regions (regions composed solely of GPP and GPO triplets), and aid in alignment and assembly of collagen molecules. The observation that individual collagen molecules are unstable at body temperature [18] could also be related to the lability of imino-poor regions, as they are less stable than imino-rich regions. Aside from affecting the global stability of the collagen molecule, subtle conformational behavior of these labile regions appears to be related to collagen turnover. Specific collagen cleavage occurs at the G-X bond in the Gly-[Ile or Leu]-[Ala or Leu] sequence, which is located upstream in an imino-poor region [19]. A detailed comparison of these regions in different collagens suggested that local conformation of collagen downstream to this cleavage site, not the local sequence, is a major factor in directing its proteolysis [19]. Structural evidence indicates that the collagen cleavage domain of a matrix metalloproteinase (MMP) can

accommodate only one α chain [20], which can occur through unwinding of the loosely wound imino-poor region.

For developing further insights into the conformational behavior of the imino-poor region and the stabilizing role of water, molecular dynamics (MD) simulation is an effective approach: it can provide atomistic details of conformational fluctuations, organization of water around collagen, and hydrogen bonding events at physiological temperatures. However, previous simulations on collagen either focused only on the hydration of imino-rich regions [21, 22, 23, 24], or studied externally induced denaturation of the imino-poor region without considering the hydration effect [25]. As mentioned above, rather than large conformational perturbations, a subtler unwinding of the imino-poor region is a biologically more relevant motion.

In this study, we perform explicit water MD simulations on collagen mimetic peptides to elucidate the dynamic role of water bridges for stabilizing the imino-poor region, as well as characterize the hydration structure along the length of the molecule. We found that individual water bridges form and break on the order of a few picoseconds, yet they are necessary to keep the imino-poor region from unwinding. On the other hand, water molecules that form a single hydrogen bond to the backbone destabilizes the triple helix as they merely provide thermal energy to the structure. In addition to surface-bound waters, we characterize the hydration shell covering the entire triple helix. The diameter of the cylinder formed by the boundary of the first hydration shell is about 14 Å, in close agreement with previous measurements of inter-collagen distances [26, 27, 28]. While the imino-poor region unwound during simulation at 300 K, the hydration shell was stable and was nearly intact even at 330 K. Water thus plays a dual role: at the level of individual molecules, it binds to the imino-poor region and controls its unwinding, while at the mesoscale level, it forms a hydration shell and acts as a biological lubricant that mediates the alignment and

self-assembly of collagen molecules.

CHAPTER II

METHODS

A. Simulation of 1BKV

We used CHARMM [29] version 31b1 with the param22 all-atom force field [30]; additional parameters for Hyp were added from an earlier study [31]. To the crystal structure 1BKV (resolution: 2.0 Å) [3], hydrogens were added using the HBUILD facility in CHARMM [32].

The peptide was solvated in a water box (dimensions: $112 \times 37 \times 37$ Å³) containing pre-equilibrated 5544 TIP3 water molecules (density: 1080 kg/m³). Periodic boundary conditions were imposed around the solvation box. Water molecules whose oxygen atoms were within a radius of 2.2 Å from heavy atoms in the peptide were deleted. After solvation, there were 5075 water molecules in the simulation box. We used a systematic energy minimization scheme to eliminate close contacts and geometric strains as described below [17]. The peptide was initially fixed and water molecules were minimized for 100 steps by steepest descent (SD) and for 400 steps by adapted basis Newton-Raphson (ABNR) method. As the next step, only the backbone was fixed and side chains along with water molecules were minimized chain by chain for 200 steps and then together for 500 steps by ABNR. The side chains were minimized chain by chain to solvate them better and to eliminate close sidechain-sidechain contacts. Finally, all constraints were removed and the system was minimized for 200 steps using ABNR.

The system was heated from 0 K, at a heating rate of 10 K/ps, and equilibrated at target temperatures ($T = 273$ K, 300 K, and 330 K) for 35 ps, with all atoms of 1BKV harmonically constrained to their original positions. After heating, equi-

Table I. Summary of simulations performed. In simulations with no charge on select backbone H and/or O atoms, water bridge formation is prevented. At 330 K, k is the spring constant of the harmonic restraint used to enforce water bridges ($1 \text{ kcal}/(\text{mol } \text{\AA}^2) = 695 \text{ pN/nm}$). The reflecting boundary prevents the peptide from diffusing out of the simulation box in longer 5 ns simulations.

Simulation Description	Peptide (Time)
At 273 K	
No constraint	1BKV (2 ns)
No charge on N-H of X	1BKV (2 ns)
No charge on N-H of X and C=O of G	1BKV (2 ns)
At 300 K	
No constraint	1BKV (2 ns)
Reflecting boundary	1BKV (5 ns)
No constraint	GPP (1 ns)
No constraint	GPO (1 ns)
No charge on N-H of X	1BKV (2 ns)
No charge on N-H of X and C=O of G	1BKV (2 ns)
At 330 K	
No constraint	1BKV (2 ns)
Reflecting boundary	1BKV (5 ns)
Reflecting boundary and harmonic constraint on water bridges $k = 10.00 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$	1BKV (5 ns)

bration followed during which temperatures were rescaled if they deviated by more than ± 2 K from target temperatures. The final production run was performed using the Leap-Frog Verlet integration algorithm with a time step of 1 fs at each target temperature and all harmonic constraints removed. Durations of production runs are shown in **Table I**. Coordinates were saved every 0.5 ps. The non-bonded pair and image atom lists were updated at each simulation step. A 12 Å distance cutoff was used for non-bonded interaction energies. Trajectories were stable during production runs with relative root-mean-square fluctuations of temperature and energy less than 0.65% and 0.14%, respectively.

At 273 K, the peptide stayed within the simulation box during the 2-ns production run. At 300 K and 330 K, after 1 ns the peptide moved, tilted, and its ends went outside the simulation box. We thus introduced a harmonic constraint, which would be applied on atoms of the triple helix only when they move outside a cylinder of diameter 34 Å about the longer axis through the center of the simulation box. In comparison, the diameter of a dehydrated collagen molecule including side chains is ≈ 12 Å [27]. The main function of the constraint was to prevent the molecule from diffusing out of the simulation box, and it had little effect on the conformation of the molecule other than slowing down of unwinding caused indirectly by its limited bending motion. In this scheme we performed 5 ns simulations at 300 K and 330 K. Each nanosecond of MD simulation took ≈ 2 days with 10 Intel Xeon (3.0 GHz) processors running in parallel.

B. GPP and GPO peptides

Structure of the GPP peptide where each α chain has the sequence (GPP)₁₀ (Protein Data Bank (PDB) ID: 1K6F, resolution: 1.3 Å) [33] was solvated in a water box

as described for 1BKV. After solvation the simulation box contained 5104 water molecules.

Crystal structure of the (GPO)₁₀ peptide is not available. We thus constructed the backbone of the peptide using the THeBuScr program [34]. Side chains were added using existing CHARMM residue topology files and energy minimization was performed residue by residue. As a test, we constructed one of the α chains of 1BKV by the above method and compared the root mean square deviation of heavy atoms from the 1BKV crystal structure, which was 1.8 Å, less than the resolution of 1BKV (2.0 Å). After solvating the GPO peptide, there were 5076 water molecules. Both systems were energy minimized, heated, and equilibrated, as done for 1BKV, and the production run at 300 K lasted for 1 ns.

C. Characterization of unwinding

As unwinding of collagen is a delicate conformational change often referred to as microunfolding [19], a sensitive method is required to characterize it. We applied the method based on local triads developed in our previous study [17]. Briefly, triangles were constructed joining the C_α atoms of Gly, X, and Y of adjacent chains along the length of the peptide (**Fig. 2**). The centroid of each triangle was chosen as the local coordinate origin and \hat{e}_1 was defined as the unit vector along the line joining the centroid to the midpoint of the line segment joining C_α atoms of chain-1 and chain-2. Setting \hat{e}_3 perpendicular to the plane fixes $\hat{e}_2 = \hat{e}_3 \times \hat{e}_1$. The first two triads and the last two triads were omitted in the analysis to reduce end effects. In this way, triads 1-4 and 15-24 form imino-rich regions, and triads 5-14 form the imino-poor region. By monitoring the Euler angle difference between two triads, the helical or torsional twist across any length of the peptide can be calculated. Mathematical details of this

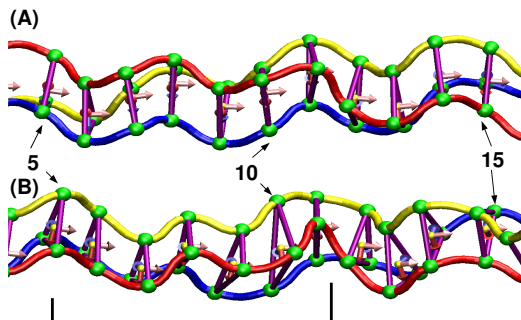


Fig. 2. Characterization of unwinding using triads. Each triad is based on the triangle that connects neighboring C_α atoms (represented by spheres) forming the cross section of the molecule. A set of arrows attached to each triangle represent the local coordinate basis of the triad, whose torsional angle can be used to quantify helical twist and unwinding [17]. Conformations are shown: (A) At start and (B) after 5 ns of MD at 330 K. Triads 5, 10, and 15 are numbered. The unwound region is marked in (B).

method can be found in Appendix A [17].

D. Water bridging and pertinent simulations

A water bridge is considered to be present only when the water oxygen atom and at least one of its hydrogen atoms are simultaneously hydrogen bonded to a bridging site as shown in Figure 1. We used a cutoff distance of 2.4 Å [35] for identifying hydrogen bond formation and checked for water bridges every 0.5 ps (the coordinate saving frequency).

For calculating the radial distribution function [See **Fig. 11(D-F)** on p.25], the distance of each water oxygen to the nearest triad on 1BKV was first identified. To eliminate the end effect, water molecules were ignored whose closest backbone atoms were at the N- or C-terminal regions outside of triads 1-24. Distances were calculated for every water molecule at each time frame during the production run. This was

computationally expensive: for every 1 ns, it took ≈ 2 days to calculate distances using 2 Intel Xeon (3.0 GHz) processors running in parallel. A histogram of water distances was made by counting the total number of water molecules during the simulation at a radial distance r , with a bin size $\Delta r = 0.025$ Å. Water molecules that belong to the same bin in the histogram form a contour that follows the periphery of three circles of radius r , respectively centered at the three C_α atoms in the corresponding triad. The contour length s was calculated analytically by assuming that each triad is an equilateral triangle of the edge length 4.75 Å. The volume element ΔV for this contour is then $\Delta V = sl\Delta r$, where l is the axial length of the region of interest. If the region of interest (imino-poor, imino-rich, or overall) contains N triads, $l = N\Delta l$, where $\Delta l = 3.65$ Å, is the average distance between centroids of triads. Finally, the density of water oxygen at r was calculated by dividing the number of water molecules in the histogram by $\Delta V \cdot (\text{number of time frames})$.

In Figure 11(D,E), water density at $r = 7$ Å in the imino-rich region is lower than that in the imino-poor region. This is because we did not compensate for the absence of water outside of the simulation box, and the tilting motion of the peptide rendered ends of the peptide (imino-rich regions) closer to the simulation boundary. On the other hand, the density at $r = 7$ Å in the imino-poor region (0.041 molecules/Å³) is slightly higher than the bulk water density under standard condition, 0.036 molecules/Å³ (= 1.08 g/ml). This is because of the use of a uniform $\Delta l = 3.65$ Å, which in fact should be larger due to the looser twist of the region. However, these factors have no effect on our main conclusion since only locations of peaks and voids in the radial distribution function that are close to 1BVK (< 3.0 Å), are important. For hydration in the imino-rich domain, these locations are consistent with a previous MD simulation [21].

CHAPTER III

RESULTS

For analyzing the behavior of the imino-poor region, we used three different mimetic peptides: 1BKV, GPP and GPO peptides. Each α chain of 1BKV has (POG)₃ and (POG)₄ respectively at the N- and the C-terminus, with a 9-residue long imino-poor domain in between (**Fig. 1**). The sequence is derived from a region of type III collagen that is important in MMP binding and cleavage. For comparison, we also tested peptides composed solely of GPP and GPO triplets. To support our conclusion, we performed various sets of simulations at three different temperatures (**Table I**).

A. Imino-poor region is more prone to unwinding

Unwinding of 1BKV can be quantified by measuring changes in torsional angles of individual triads (**Fig. 2**; see Methods) [17]. At 273 K, no noticeable unwinding occurred during the first 2 ns, while it was more evident at 300 K and 330 K (**Fig. 3**). At 300 K, unwinding was limited between triads 5-9 within the simulation time, while it further propagated at 330 K. The 5-ns simulations in **Fig. 3(B,C)** were performed in the presence of a cylindrical reflecting boundary to prevent the peptide from diffusing out of the simulation box (see Methods; Table I), which slowed down the conformational fluctuation of collagen by suppressing its global bending motion. Simulations performed for 2 ns at 300 K and 330 K without the reflecting boundary showed faster unwinding, but one end of the peptide moved out of the simulation box (data not shown). If a larger simulation box without the reflecting boundary were used (computationally more expensive), it would only accelerate unwinding for a given simulation time. Regardless of variations in simulation conditions, it is clear that the imino-poor region is more likely to unwind compared to imino-rich regions.

As previously observed [17], unwinding initiates at the Gly-Ile bond (triads 5-8), which is a major bond cleaved by collagenases [19]. Its spontaneous unwinding would facilitate cleavage since collagenases cleave one α chain at a time [20].

As a comparison, we have performed the same simulation using triple helices where individual α chains were composed of (GPP)₁₀ or (GPO)₁₀. No unwinding was observed at 300 K during the 1-ns simulation time (**Fig. 4**). Note that experimentally measured melting temperatures of GPO (≥ 323 K) and GPP (≈ 307 K) peptides are higher than that of 1BKV (≈ 290 K) [36, 37]. The inductive effect of Hyp in the Y position would further add to the structural stability of imino-rich regions compared to Pro in the Y position, which is why the GPO peptide is more stable than the GPP peptide [36]. The force field used for the MD simulation does not incorporate the inductive effect, thus the present simulation cannot accurately account for the difference in stability of the GPP and GPO peptides. However, this has no effect on the observed unwinding of the imino-poor region.

B. Dynamics of surface water bridges

There are nine water bridging sites in the imino-poor domain of 1BKV (**Fig. 1**). Although multiple water molecules can join by hydrogen bonds to form a bridge [26], in the present study, we analyzed only bridges mediated by a single water molecule between backbones of α chains. As explained below, hydrogen bonds dynamically form and break, which would make bridges involving multiple water molecules to play a less stabilizing role compared to those involving a single water molecule.

Although we did not keep crystal water bridges, they re-formed soon after MD started. We measured two different characteristic times to study the nature of water bridges. The lifetime of a given water bridge is the duration for which a water

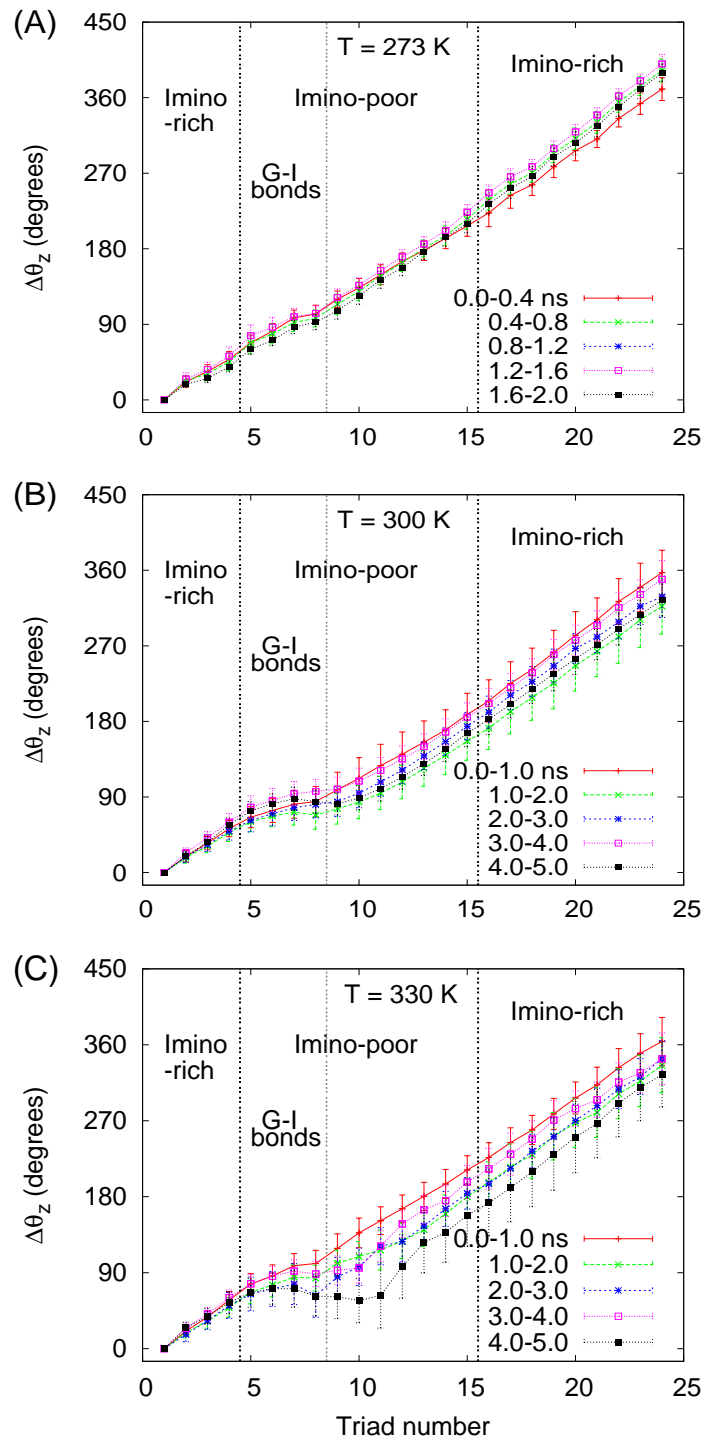


Fig. 3. Temperature-dependent unwinding. The local slope of each curve is directly proportional to the helicity at the corresponding triad. Each point corresponds to the helical angle difference from the first triad averaged over successive 0.4-ns or 1-ns intervals (\pm standard deviation). Triads 5-8 contain Gly-Ile bonds, where unwinding initiates.

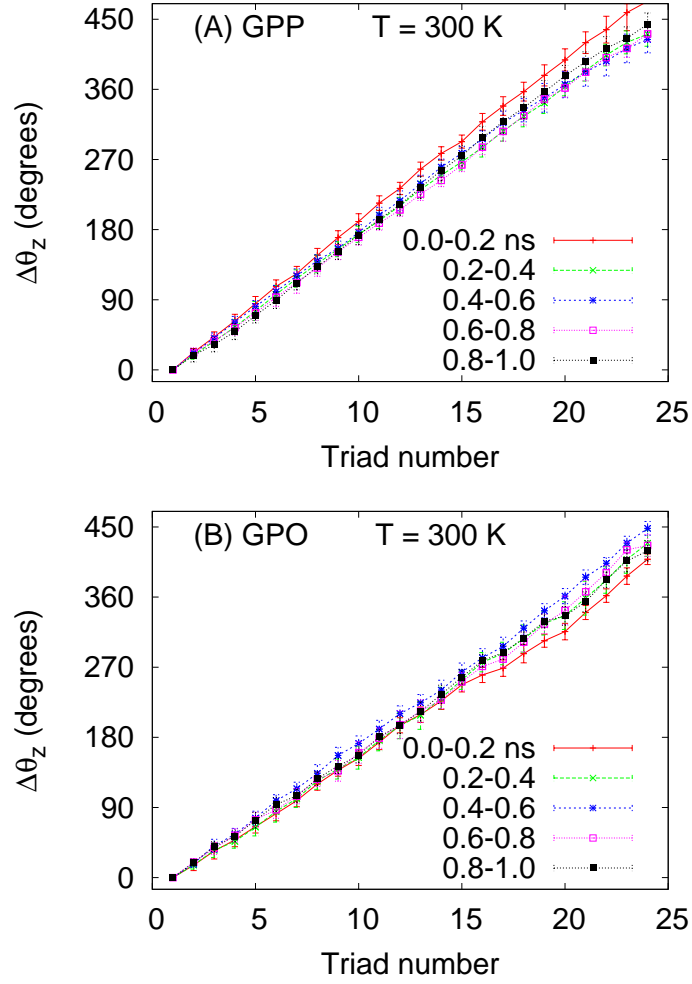


Fig. 4. Torsional map of imino-rich peptides at 300 K (*see*, **Fig. 3**). (A) (GPP)₁₀ and (B) (GPO)₁₀. Note that $\Delta\theta_z$ at triad 24 is larger than those in **Fig. 3**, due to their tighter twist.

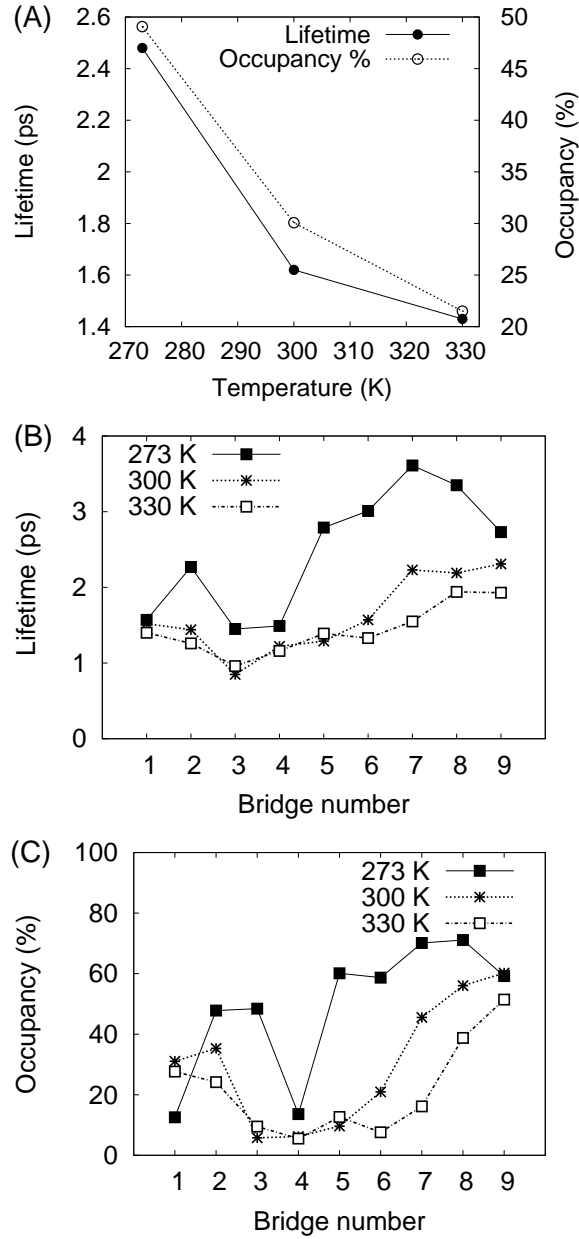


Fig. 5. Analysis of water bridging events. (A) Average lifetime and occupancy (occupancy time divided by the simulation time) of water bridges. Standard deviations in lifetimes are 3.03 ps (273 K), 1.74 ps (300 K), and 1.44 ps (330 K). Note that standard deviations are larger than averages, indicating that the distribution of lifetimes is asymmetric with a long tail at large lifetimes. (B) Average lifetimes and (C) occupancy of individual water bridging sites (*cf.*, Fig. 1).

bridge stays without breaking either of its hydrogen bonds to the peptide backbone. The occupancy time is the total time for which a water bridging site is occupied during simulation (sum of individual lifetimes) [**Fig. 5(A)**]. The average lifetime of water bridges decreases with increasing temperature. More importantly, lifetimes were on the order of a few picoseconds, which is less than that of individual water-to-water hydrogen bonds in a bulk water [38]. Average occupancy also decreases with increasing temperature. Note that we did not count water bridging events shorter than the frequency of saving coordinates during MD, which was 0.5 ps. If these events were considered, the occupancy would increase even more. Frequent formation of water bridges with short lifetimes indicates their dynamic nature, so they should not be considered to be in an ice-like phase. One of the main arguments against water bridges playing a stabilizing role was the high entropic cost of holding water molecules [9]. However, their short lifetimes suggest that there would be little entropic cost associated with water bridging events, which is believed to be generally true for surface-bound water molecules [39].

To further elucidate water bridge formation, lifetime and occupancy of the nine water bridges were individually measured [**Fig. 5(B,C)**]. Overall, bridges near the Gly-Ile region were shorter-lived with a lower occupancy, compared to those near the C-terminal end of the imino-poor domain. B1 and B2 in the Gly-Ile region were slightly more active than B3-B5, probably due to the staggered arrangement of α chains, which places the GPO region next to the initial Gly-Ile bonds. As mentioned above, Gly-Ile and Gly-Leu are the bonds cleaved by collagenases, at about three-fourths of the length of collagen from the N-terminus [40]. There are many Gly-Ile and Gly-Leu bonds in collagen, but the specificity of collagenases depends on the labile imino-poor domain downstream of the cleavage site [19]. Consistent with this, we found that Gly-Ile region and the bonds immediately downstream, where unwinding

initiates and propagates, to have the weakest water bridges, which can be seen more clearly by monitoring individual water bridges over time (**Fig. 6**).

C. Loss of water bridges causes unwinding

Above observations suggest that loss of water bridges at higher temperatures could be responsible for unwinding of the region. We measured the number of water bridges and the helical angle across the imino-poor region of 1BKV as a function of time (**Fig. 7**). No correlation was noticeable at 273 K, as the region did not unwind. At 300 K, the helicity of the imino-poor domain fluctuates more compared to 273 K, as seen during 0.5-2.0 ns and 3.5-5.0 ns intervals. The number of water bridges correlates well with this behavior. The correlation is also seen at 330 K, where the helical angle fluctuates more and unwinding is evident.

As the number of water bridging events correlates well with unwinding of the imino-poor domain, we tested if continuous presence of water bridges prevents unwinding. We performed a 5-ns simulation at 330 K during which the nine water bridges were maintained by harmonic constraints. The constraint was activated only when hydrogen or oxygen of each water forming a bridge were more than 2.4 Å away from respective backbone atoms in the peptide, a commonly used cutoff distance for identifying hydrogen bonds [35]. Thus no extra force was applied when water bridges were formed, and harmonic constraints were used only to restore water bridges when they were broken. The spring constant of the constraint was 10 kcal/(mol·Å²). As before, a cylindrical reflecting boundary was applied to prevent the peptide from moving outside the simulation box (Table I). During the simulation run, the average occupancy of water bridges was 79.3%, higher than that at 273 K without the constraint [**Fig. 5(A)**]. Despite the simulation temperature of 330 K, the helicity of

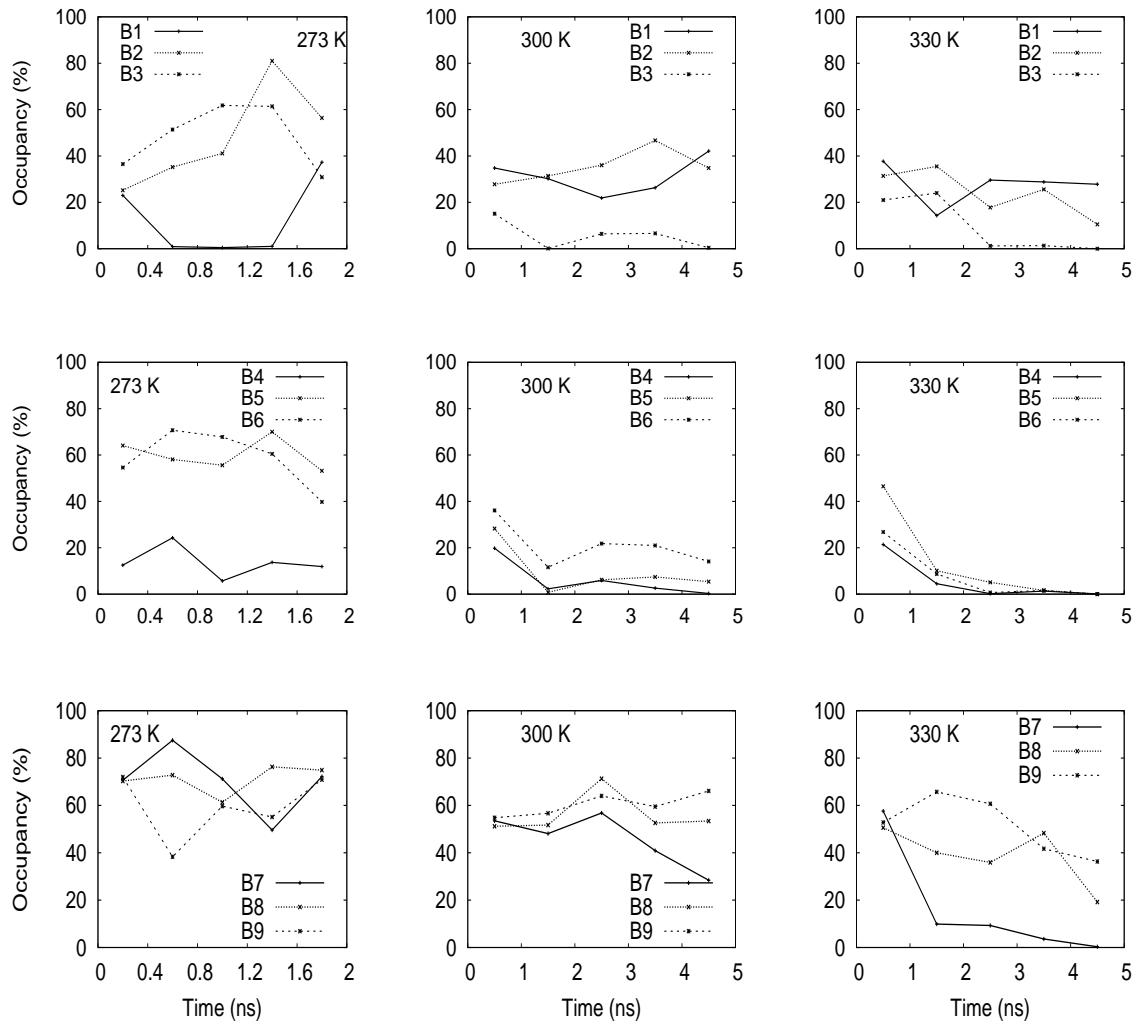


Fig. 6. Average occupancy of individual water bridges. The total simulation time at 273 K (2 ns), 300 K (5 ns), and 330 K (5 ns) was divided into five intervals as shown in the graph. Bridge numbers B1-B9 are from **Fig. 1**.

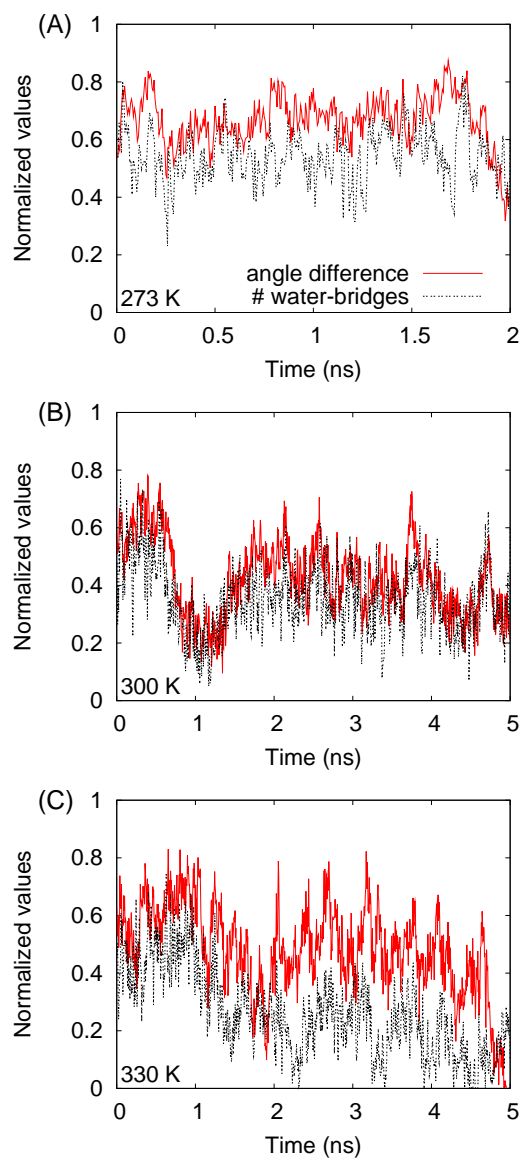


Fig. 7. Correlation between the number of water bridges and the helical angle difference across the imino-poor region (triads 5-15). The angle difference was normalized within the 20° - 180° window and the number of water bridges was normalized within the 0-9 window. (A) 273 K, (B) 300 K, and (C) 330 K

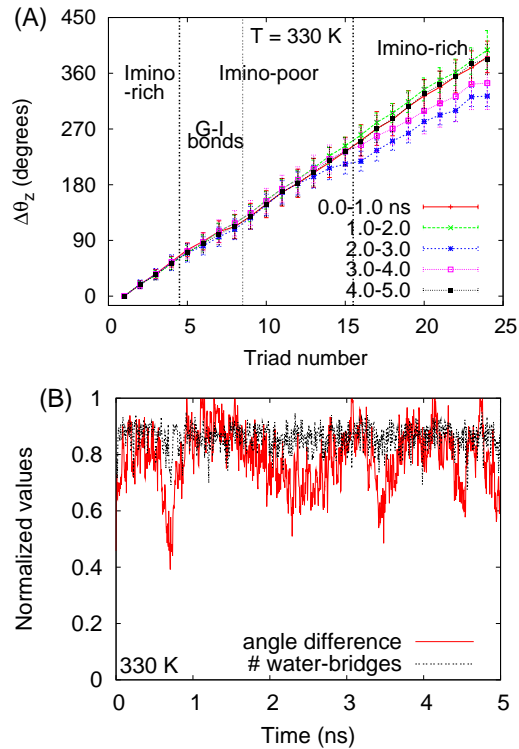


Fig. 8. Suppression of unwinding by enforcing water bridge formation. (A) Torsional map (*see*, **Fig. 3**). (B) Number of water bridges and the helical angle difference across the imino-poor region (triads 5-15; *see*, **Fig. 7**).

the molecule remained mostly intact. Only during the 2-3 ns period, triads 14-15 unwound slightly, which subsequently re-wound, restoring the original helicity by the 4-5 ns period [**Fig. 8(A)**]. Triad 14 contains the Gly-Leu bond (**Fig. 1**). As explained in Discussion, together with the Gly-Ile bond, it is also a common cleavage site and has a stronger tendency to unwind than other bonds due to the high β -sheet propensity and hydrophobicity.

D. Dual role of water as stabilizer and as thermal agitator

If water bridges are mainly responsible for stabilizing the imino-poor region, preventing water bridge formation should destabilize the region. We tested this by selectively nullifying point charges on the backbone atoms that form hydrogen bonds with water. Two cases were considered:

Case 1. No charge on N-H of X in the imino-poor region [**Fig. 9(B)**].

Case 2. No charge on N-H of X in the imino-poor region and on C=O of Gly [**Fig. 9(C)**].

In Case 1, although water bridges cannot form, a single hydrogen bond can form with water via C=O of Gly, while in Case 2, no water hydrogen bond can form on bridging sites or on Gly along the entire peptide. As expected, in Case 1 the molecule unwound even at 273 K [**Fig. 10(A)**]. Surprisingly, in Case 2, the molecule did not unwind until the end of the 2-ns simulation [**Fig. 10(B)**]. Water can supply thermal agitation directly to the peptide backbone through hydrogen bonding to the backbone atoms, possibly via its pivoting motion. In Case 2, there are no hydrogen bonds between the backbone and water, resulting in weaker thermal agitation, which likely explains the absence of unwinding. The molecule in Case 2 can thus be considered to be in a neutral state in which there is neither a stabilizing water bridge nor strong enough thermal agitation, which possibly caused the torsional map to be less smooth compared to other cases (**Fig.10**). However, when we did the same test at 300 K, thermal energy was enough to unwind the peptide in Case 2 within 1 ns (data not shown). Taken together, the above series of simulations confirm that thermal unwinding of the imino-poor region is caused by loss of water bridges.

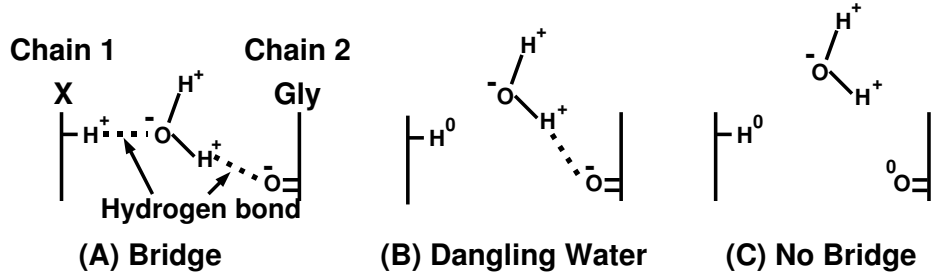


Fig. 9. Preventing water bridge formation. A bridging site with amino hydrogen and carboxylic oxygen from two adjacent chains along with a water molecule is shown. Signs of point charges (+, −, or 0) are marked above respective atoms. (A) Normal case when a water bridge is formed. (B) The point charge on N-H of X in the imino-poor region is removed, preventing bridge formation. But water can form a hydrogen bond with C=O of Gly (Case 1). (C) In addition to (B), charge on C=O of Gly is removed (Case 2).

E. Organization of hydration layers depends on the imino-acid content

Extending the analysis of water bridges, we considered spatial organization of water around the entire triple helix. We first calculated the radial distance of water oxygen atoms from the backbone. We chose the backbone instead of the helical axis as the reference for distance measurement, since the helical axis is not well-defined during MD simulations due to the conformational fluctuation of the molecule. For clarity, we grouped backbone atoms belonging to the same triad and plotted the distance of each water oxygen from the closest triad. Example snapshots are shown in **Fig. 11(A-C)**, which reveal that water oxygens are closer to the backbone in the imino-poor region. On the other hand, a similar plot for water hydrogens shows that they are closer to the backbone throughout the molecule (**Fig. 12**).

To further understand these differences, we calculated the radial distribution function of water oxygen in imino-rich and imino-poor regions, separately and to-

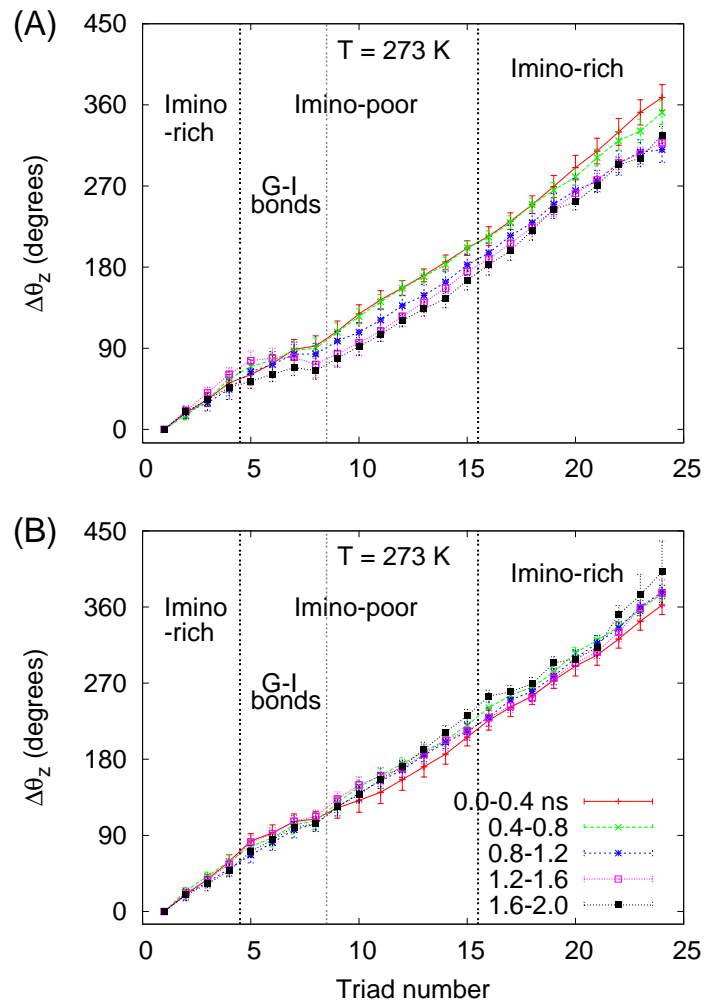


Fig. 10. Torsional map of 1BKV in the absence of water bridges. (*cf.*, **Fig. 3**). (A) Case 1: No charge on N-H of X. Unwinding was observed even at 273 K. (B) Case 2: No charge on N-H of X and C=O of Gly. Unwinding was not observed during the simulation.

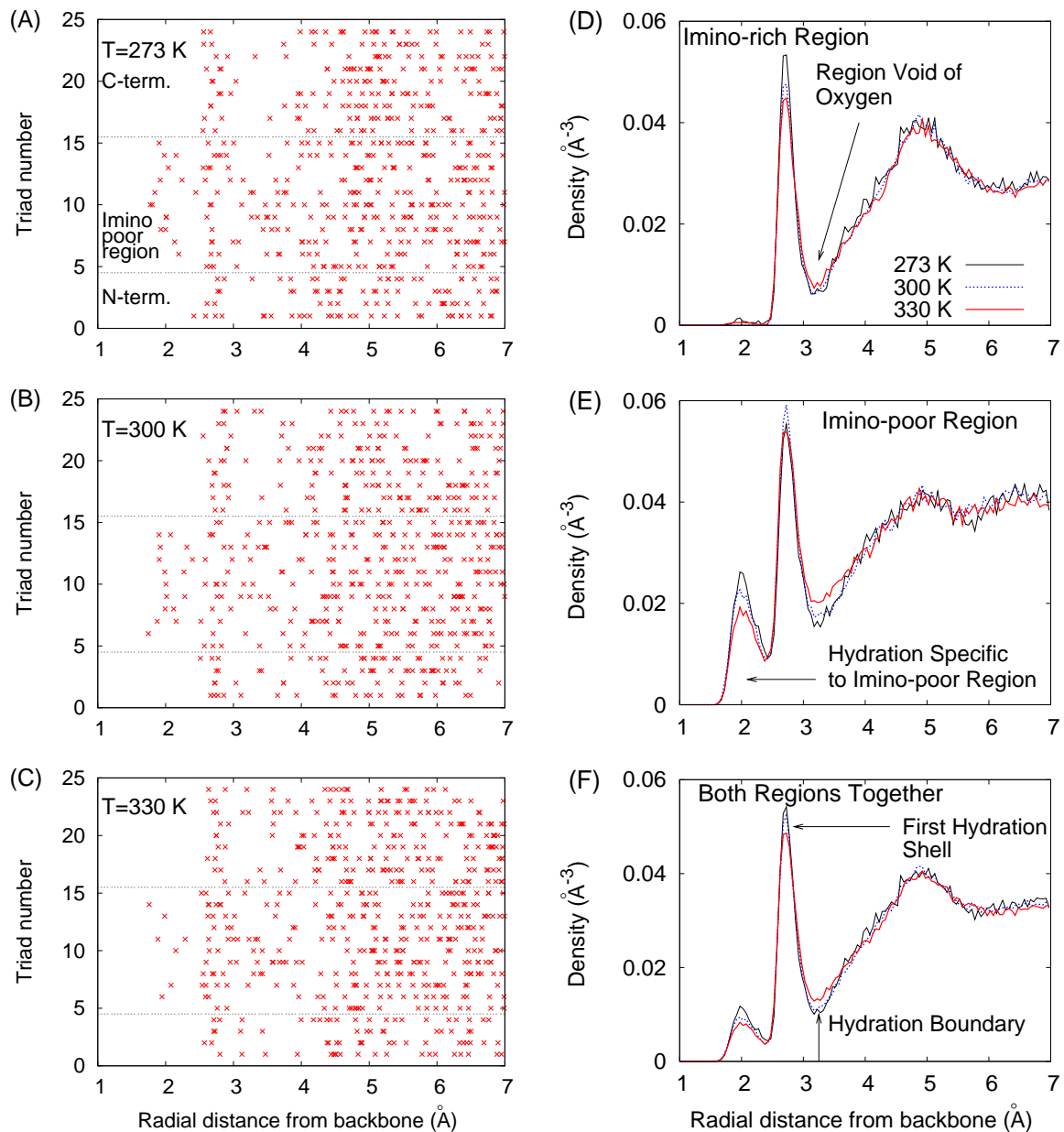


Fig. 11. Hydration structure around the molecule. (A-C) Snapshot of water oxygens (at 970 ps; an arbitrarily chosen time) at (A) 273 K, (B) 300 K, and (C) 330 K. Water oxygens are closer to the backbone in the imino-poor region compared to imino-rich regions. (D-F) Average radial distribution of water oxygen atoms over the simulation period (273K: 2ns; 300,330 K: 5ns). The overall profile is preserved during the simulation, although the peak height decreases slightly as temperature increases (see also **Fig. 14**). (D) Imino-rich region, (E) imino-poor region, and (F) both regions together.

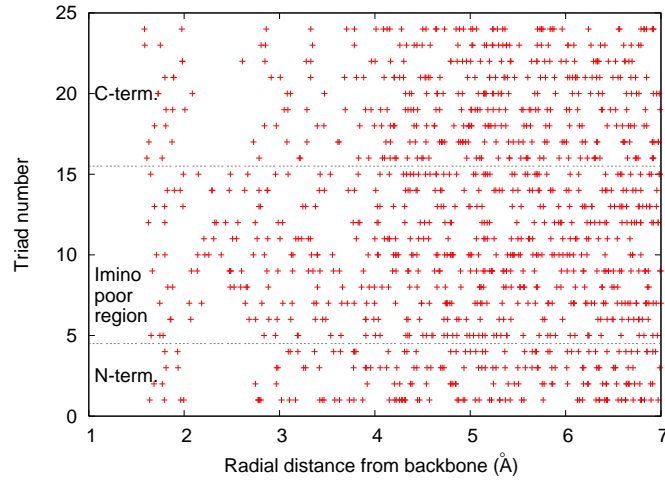


Fig. 12. Position of water hydrogens around collagen at 970 ps at 273 K. Note that the first shell of water hydrogens ($\leq 2.00\text{\AA}$) are closer to the backbone than water oxygens in the imino-poor region [Fig. 11(A-C)]. This reflects the shorter length of the hydrogen bond between water hydrogen and backbone oxygen compared to that between water oxygen and backbone hydrogen, due to the stronger electrostatic interaction (a larger backbone point charge) in the former case.

gether [Fig. 11(D-F)]. The most prominent difference between the two regions is that there are two peaks in the imino-poor region in contrast to one in the imino-rich region. The first peak in the imino-poor region is a spatially discontinuous shell of water molecules that include bridges and those hydrogen bonded to N-H of X. Indeed the first peak is at ≈ 2.0 Å from the backbone, that corresponds to the typical hydrogen bonding distance between water oxygen and backbone hydrogen. Imino-rich regions lack the backbone amide hydrogen (N-H of Gly is used for backbone hydrogen bonding; Fig. 1), and are helically more tightly wound than imino-poor regions [3]. Side chain imino rings sterically hinder the entry of oxygen near the backbone. In contrast, hydrogen atoms, being small and carrying less charge, can be close to the backbone in all regions of the molecule (Fig. 12). It should be noted that, if the distance of water oxygen is measured from the entire collagen molecule including side chains, a low density (0.01 Å⁻³) first peak can be seen even for a GPO triple helix [21]. However, this peak is caused by water molecules hydrogen bonded discretely to the -OH group of Hyp, thus cannot be regarded as a semi-continuous shell covering the molecule. Furthermore, their distance from the backbone is larger (similarly from the helical axis), thus calculating distances based on the peptide backbone instead of the entire molecule provides a clearer description of the radial organization of water around the triple helix.

The second peak is at ≈ 2.8 Å, and is non-specific to the imino acid content. At this distance, water oxygens cannot form hydrogen bonds with backbone atoms. Instead, water hydrogens form bonds with C=O of the backbone throughout the molecule. This coincides with the first hydration shell identified in the x-ray structure of the GPP peptide [28]. A comparison between distributions of water oxygens (Fig. 11) and water hydrogens (Fig. 12) reveals a subtle difference between the first and the second peaks in Fig. 11 regarding orientations of water molecules. In the

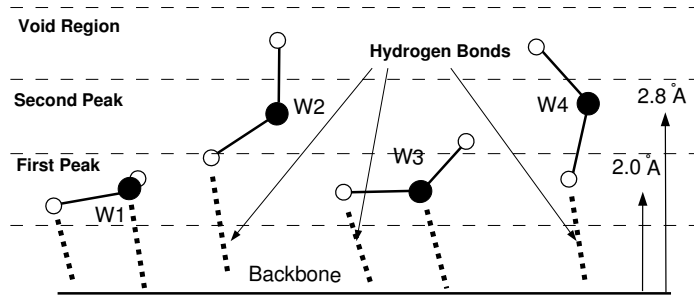


Fig. 13. A cartoon representing the orientation of water molecules (W1–W4; not to scale). Black circle: oxygen. White circle: hydrogen. Dashed lines (– –) separate important hydration regions including the two peaks and the void region in **Fig. 11(D-F)**. W1 and W3: Typical orientations of water bridges found only in the imino-poor region, where one O–H bond is approximately parallel to the surface of collagen. W2 and W4: Typical orientations radially perpendicular to the backbone, which are present in both imino-poor and imino-rich regions. Vertical arrows show approximate distances of hydration peaks from the backbone of 1BKV. Note that only atoms that lie within the first peak (≤ 2.4 Å) can form hydrogen bonds with the backbone.

first peak, the O–H bond in water forming a bridge is parallel to the surface of the triple helix (**Fig. 13**; W1,W3) such that oxygen as well as one hydrogen of water can be hydrogen bonded to the backbone. In the second peak the O–H bond of water molecule is oriented radially perpendicular to the backbone (**Fig. 13**; W2,W4). Such difference in the orientation of water between the two peaks leads to regions devoid of water oxygens at ≈ 2.40 Å and ≈ 3.25 Å, which are also visible in **Fig. 11(A-C)**. The void region after the second peak is more pronounced in the imino-rich region than in the imino-poor region due to the orientational variation of water molecules in the first peak in the imino-poor region, (*e.g.*, **Fig. 13**; W1 *vs.* W3) which disrupts the order of water molecules in the second peak.

F. The first hydration shell defines the boundary for collagen self-assembly

The density of water in the first peak is about half the bulk water density seen at around 7 Å from the backbone, as seen in the imino-poor region [**Fig. 11(E)**]. The second peak is higher than the bulk density [**Fig. 11(D-F)**], suggesting a densely packed hydration structure. The radially parallel orientation of water (**Fig. 13**; W2, W4) in the second peak suggests that they may assist in packing water molecules in the space between side chains. Such packing could be entropically favored, similar to the entropy-driven filling of protein cavities by water, as noted previously [41, 42, 43].

At higher temperatures, heights of the peaks vary slightly, but the overall structure of the radial distribution function does not change significantly [**Fig. 11(D-F)**]. The first peak decreases with increasing temperature, as less water is available for bridging at elevated temperatures. The second peak (first hydration shell) is preserved despite unwinding of the imino-poor region (**Fig. 14**). The void region also retains its low water oxygen density at all temperatures. These results suggest that, when two collagen molecules with similar hydration structures self-assemble, they would do so at the boundary of their first hydration shell such that void regions overlap. It may not be the exact boundary every time two collagen molecules self-assemble, but would set the minimum distance between the two owing to the stability of the hydration shell. Backbone atoms in collagen are about 3-4 Å from the helical axis of the molecule [34]. Adding this distance to the location of the void after the second peak gives 6.25-7.25 Å from the helical axis, making the inter-axial spacing when two helices self-assemble to be 12.5-14.5 Å. This value agrees well with the 13-14 Å axis-to-axis distance measured in x-ray structures [26, 28, 44].

We also followed the time evolution of the radial distribution of water oxygens (**Fig. 14**). The hydration layer in the imino-rich region fluctuates less than that in the

imino-poor region. For example, at 300 K, the peak density of the radial distribution function in imino-rich regions fluctuated by 42.8%, compared to 49.7% (86.9%) for the second (first) peak in the imino-poor region. This data shows that imino-rich regions are not only important in imparting structural stability, but they also define a clearer hydration boundary for self-assembly. Disturbance of the radial distribution function in the imino-poor region is particularly pronounced at 330 K (**Fig.** 14). Such fluctuating nature of the hydration layer in the imino-poor region would make it easier for enzymes or FACITs (fibril associated collagens with interrupted triple helices) to bind, since a well ordered layer of water, as found in the imino-rich region, would impose a stronger repulsive hydration force.

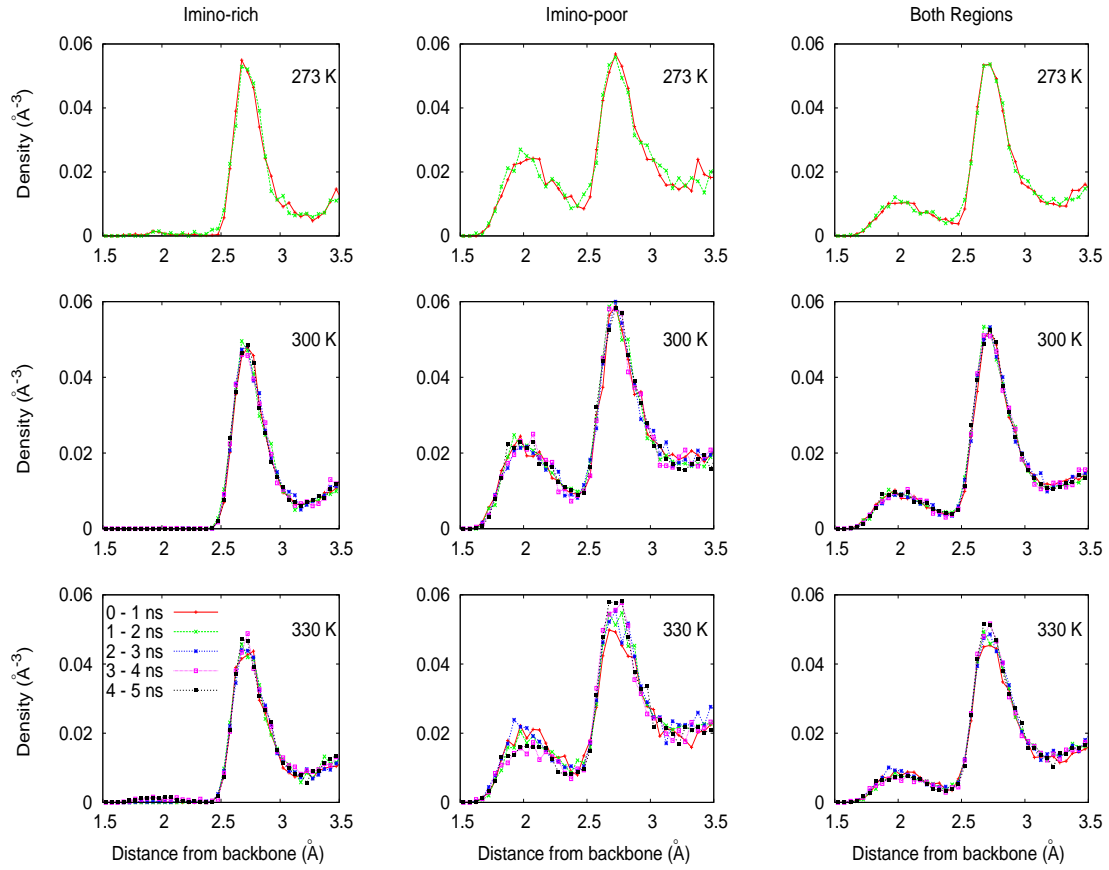


Fig. 14. Average radial distribution function of water oxygens at different times and temperatures. Column 1: imino-rich, column 2: imino-poor, column 3: both regions together.

CHAPTER IV

DISCUSSION AND CONCLUSION

A. Discussion

In the present study, we elucidate the linkage between the conformational behavior of collagen and the hydration dynamics. A comparison between 1BKV and GPP or GPO peptides reveals that the imino-poor region is much more likely to unwind than imino-rich regions. In 1BKV, unwinding initiated at the Gly-Ile region, a typical bond cleaved by collagenases. The high β -sheet propensity of Ile [45] would prefer an extended conformation and assist with unwinding. Its hydrophobic nature could be another feature that promotes unwinding by reducing water bridge formation [Fig. 5(B,C)]. A similar behavior is expected for the other cleavage site, the Gly-Leu bond. For globular proteins, the breakage of internal hydrophobic contacts is a key event in unfolding [46]. As collagen is a linear molecule where hydrophobic side chains are inherently exposed to water, imino-poor regions that contain hydrophobic residues, including those of collagen cleavage sites, may serve as local nucleation points for micro-unfolding of collagen, a behavior thought to be important for collagen proteolysis [19].

Single water bridges between α chains in the imino-poor region [3] are highly dynamic with lifetimes on the order of only a few picoseconds. Nevertheless, they are crucial for stabilizing the imino-poor region: the helicity of the region closely followed the number of water bridges (Fig. 7). Furthermore, enforcing water bridges prevented unwinding of the helix (Fig. 8) whereas inhibiting water bridges resulted in easier unwinding (Fig. 10). In the latter case, we have also found that water not only stabilizes the structure when it forms a bridge, but when it forms a single

hydrogen bond to the backbone instead of a bridge, it destabilizes the molecule by supplying thermal energy. Such a dual role of water as a stabilizer and as an agitator is consistent with the non-monotonic dependence of protein stability on the fraction of organic solvents in water [13, 47]. In the case of type-I collagen, its denaturation temperature exhibits a U-shaped curve as a function of the alcohol content in aqueous solution, with a minimum at $\sim 30\%$ by weight [13]. The initial destabilization up to 30% is possibly due to the loss of surface bound water bridges through disruption of the first hydration shell and the first hydration peak (see below), while stabilization at higher alcohol content is due to the lack of thermal agitation by ‘dangling’ water molecules [*cf.*, **Fig. 9(B)**]. Consistent with this picture, there is a rather general report that the conformational mobility of proteins is kinetically restricted in anhydrous media [47].

Unlike specific bridges in the imino-poor region, the overall hydration around the molecule was found to be stable and little affected by unwinding or decreased lifetimes of water hydrogen bond networks at higher temperatures (**Fig. 11**). In the imino-rich region, bridges involving the -OH group of Hyp connected by multiple water molecules have been observed in crystal structures [28]. Such bridges would be part of the first hydration shell (the second peak in **Fig. 11(F)**) which spans both imino-poor and imino-rich regions. The position and geometry of this hydration layer should be maintained not by specific hydrogen bonds or water bridges, but rather by steric packing of water molecules near the surface, as illustrated in **Fig. 13**. In this case, hydrogen bonds merely control water orientations, and their temperature-dependent lifetimes do not determine the organization of hydration layers. Our view agrees with the earlier observation that water hydrogen bonds with Hyp do not have a major role in the stability of collagen [8]. Rather, backbone water bridges in the imino-poor region, as discussed above, do have a stabilizing effect on collagen.

Our results also indicate that the first hydration shell is a biological lubricant in collagen self-assembly and determines the inter-helix distance. Without the hydration shell, collagen molecules would aggregate by forming non-specific bonds, resulting in kinetically trapped amorphous states [23, 48].

Leikin and co-workers measured the osmotic stress in removing water between collagen molecules and reported a temperature independent, exponentially increasing force for inter-helix distances below $\approx 16 \text{ \AA}$ [49, 50, 51]. This distance is close to our estimate 12.5-14.5 \AA , based on the void region after the first hydration shell. As seen in **Fig. 11(D-F)**, there is a modest second hydration shell (third hydration peak), which may also contribute to the repulsive force and extend its range. Leikin and co-workers explained that the exponentially repulsive force arises due to structural changes in the hydration layer. Our analysis provides a microscopic basis for this argument. If two collagen molecules become closer than the distance that can accommodate their respective hydration shells, steric packing of water molecules (**Fig. 13**) would be disrupted along the length of collagen, which would be much more unfavorable than disrupting hydrogen bond networks. They also reported a weak, temperature dependent attractive force at distances greater than $\approx 16 \text{ \AA}$ [50]. Such a force may arise when two collagen molecules are at distances such that the oxygen devoid regions (**Fig. 11**) flank each other. The void regions contain only smaller hydrogen atoms, and filling such regions would be entropically favorable. This attraction is expected to be weaker than repulsion at shorter distances, as it would be entropic, rather than steric in origin, where the latter is temperature independent.

Attraction between collagen molecules disappears under various conditions, [51] including addition of glycerol or glucose. This is likely to be related to the disruption of the first hydration shell, as attraction or repulsion between collagen molecules are based mainly on hydration forces. Apart from collagen, short-range exponential

repulsion is common to other biomolecules such as DNA or lipids [49], possibly based on a similar origin.

Although the present analyses are based on short collagen mimetic peptides, it has implications on the conformational behavior of native collagens. The free energy barrier for unwinding would be low, as unwinding was observed in our relatively short MD simulations, which is consistent with the instability of isolated collagen molecules at body temperature [18]. Even when they are cross linked and form a more stable collagen fiber, imino-poor regions (thermally labile domains) [16] would still thermally undulate in its helicity, as there are only a few cross-linking points along the length of the molecule, especially in the early stage of fibril formation. This can happen without disrupting the alignment of collagen molecules in the fiber, since the first hydration shell was found to be intact even when the molecule unwinds. This view suggests that a mechanically *unloaded* collagenous matrix is vulnerable to degradation by collagenases, which require unwinding of the cleavage domain [20]. In contrast, mechanical load on collagen would suppress unwinding, thus causes the matrix less prone to degradation. The load-dependent cleavage of collagenous matrix has been experimentally observed [52, 53, 54, 55].

B. Conclusion

Conformational stability of collagen molecules affects their hierarchical self-assembly and turnover. Although unwinding of the imino-poor region had been suggested to be a key conformational feature related to collagen cleavage, its atomistic details have been missing. The present study elucidates the atomistic picture and illustrates that unwinding initiates locally in the imino-poor region and is spontaneous at physiological temperatures. Collagenases thus can take advantage of such innate feature

of collagen, rather than actively unwinding the molecule for cleavage. Unwinding, which is mediated by water, also provides a biological mechanism through which mechanical tension on the collagen molecule is linked to its cleavage and turnover – a molecular-level protein structure-function relation crucial for the tissue-level homeostasis. Biological roles of water as a stabilizer, an agitator, or as a lubricant, as found through the present study, have implications for higher-level collagen assemblies as well as hydration behaviors of other biomolecules in general.

REFERENCES

- [1] R. Y. Yee, S. W. Englander, and P. H. V. Hippel, "Native collagen has a two-bonded structure," *J Mol Biol*, vol. 83, pp. 1–16, 1974.
- [2] A. Rich and F. H. C. Crick, "The molecular structure of collagen," *J Mol Biol*, vol. 3, pp. 483–506, 1961.
- [3] R. Z. Kramer, J. Bella, P. Mayville, B. Brodsky, and H. M. Berman, "Sequence dependent conformational variations of collagen triple-helical structure," *Nat Struct Biol*, vol. 6, pp. 454–457, 1999.
- [4] J. Rosenbloom, M. Harsch, and S. Jimenez, "Hydroxyproline content determines the denaturation temperature of chick tendon collagen," *Arch Biochem Biophys*, vol. 158, pp. 478–484, 1973.
- [5] R. A. Berg and D. J. Prockop, "The thermal transition of a non-hydroxylated form of collagen. evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen," *Biochem Biophys Res Commun*, vol. 52, pp. 115–120, 1973.
- [6] G. N. Ramachandran, M. Bansal, and R. S. Bhatnagar, "A hypothesis on the role of hydroxyproline in stabilizing collagen structure," *Biochim Biophys Acta*, vol. 322, pp. 166–171, 1973.
- [7] J. Bella, B. Brodsky, and H. M. Berman, "Hydration structure of a collagen peptide," *Structure*, vol. 3, pp. 893–906, 1995.
- [8] J. Engel, H. T. Chen, D. J. Prockop, and H. Klump, "The triple helix in equilibrium with coil conversion of collagen-like polytripeptides in aqueous and non-aqueous solvents. Comparison of the thermodynamic parameters and the binding

- of water to (L-Pro-L-Pro-Gly) $_n$ and (L-Pro-L-Hyp-Gly) $_n$,” *Biopolymers*, vol. 16, pp. 601–622, 1977.
- [9] S. K. Holmgren, K. M. Taylor, L. E. Bretscher, and R. T. Raines, “Code for collagen’s stability deciphered,” *Nature*, vol. 392, pp. 666–667, 1998.
- [10] L. E. Bretscher, C. L. Jenkins, K. M. Taylor, M. L. DeRider, and R. T. Raines, “Conformational stability of collagen relies on a stereoelectronic effect,” *J Am Chem Soc*, vol. 123, pp. 777–778, 2001.
- [11] D. A. Slatter, C. A. Miles, and A. J. Bailey, “Asymmetry in the triple helix of collagen-like heterotrimers confirms that external bonds stabilize collagen structure,” *J Mol Biol*, vol. 329, pp. 175–183, 2003.
- [12] J. A. Ramshaw, N. K. Shah, and B. Brodsky, “Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides,” *J Struct Biol*, vol. 122, pp. 86–91, 1998.
- [13] K. Fukuda, T. Nezu, and Y. Terada, “The effects of alcoholic compounds on the stability of type I collagen studied by differential scanning calorimetry,” *Dent Mater J*, vol. 19, pp. 221–228, 2000.
- [14] R. Z. Kramer, M. G. Venugopal, J. Bella, P. Mayville, B. Brodsky, and H. M. Berman, “Staggered molecular packing in crystals of a collagen-like peptide with a single charged pair,” *J Mol Biol*, vol. 301, pp. 1191–1205, 2000.
- [15] R. Z. Kramer, J. Bella, B. Brodsky, and H. M. Berman, “The crystal and molecular structure of a collagen-like peptide with a biologically relevant sequence,” *J Mol Biol*, vol. 311, pp. 131–147, 2001.

- [16] C. A. Miles and A. J. Bailey, “Thermally labile domains in the collagen molecule,” *Micron*, vol. 32, pp. 325–332, 2001.
- [17] K. M. Ravikumar, J. D. Humphery, and W. Hwang, “Spontaneous unwinding of a labile domain in a collagen triple helix,” *J Mech Matl Struct*, vol. 2, pp. 999–1010, 2007.
- [18] E. Leikina, M. V. Merts, N. Kuznetsova, and S. Leikin, “Type I collagen is thermally unstable at body temperature,” *Proc Natl Acad Sci USA*, vol. 99, pp. 1314–1318, 2002.
- [19] G. B. Fields, “A model for interstitial collagen catabolism by mammalian collagenases,” *J Theor Biol*, vol. 153, pp. 585–602, 1991.
- [20] L. Chung, D. Dinakarpandian, N. Yoshida, J. L. Lauer-Fields, G. B. Fields, R. Visse, and H. Nagase, “Collagenase unwinds triple-helical collagen prior to peptide bond hydrolysis,” *EMBO J*, vol. 23, pp. 3020–3030, 2004.
- [21] J.-W. Handgraaf and F. Zerbetto, “Molecular dynamics study of onset of water gelation around the collagen triple helix,” *Proteins*, vol. 64, pp. 711–718, 2006.
- [22] R. J. Radmer and T. E. Klein, “Triple helical structure and stabilization of collagen-like molecules with 4(R)-hydroxyproline in the Xaa position,” *Biophys J*, vol. 90, pp. 578–588, 2006.
- [23] I. G. Mogilner, G. Ruderman, and J. R. Grigera, “Collagen stability, hydration and native state,” *J Mol Graph Model*, vol. 21, pp. 209–213, 2002.
- [24] T. E. Klein and C. C. Huang, “Computational investigations of structural changes resulting from point mutations in a collagen-like peptide,” *Biopolymers*, vol. 49, pp. 167–183, 1999.

- [25] C. M. Stultz, “Localized unfolding of collagen explains collagenase cleavage near imino-poor sites,” *J Mol Biol*, vol. 319, pp. 997–1003, 2002.
- [26] J. Bella, M. Eaton, B. Brodsky, and H. M. Berman, “Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution,” *Science*, vol. 266, pp. 75–81, 1994.
- [27] G. D. Fullerton and M. R. Amurao, “Evidence that collagen and tendon have monolayer water coverage in the native state,” *Cell Biol Intl*, vol. 30, pp. 56–65, 2006.
- [28] R. Z. Kramer, L. Vitagliano, J. Bella, R. Berisio, L. Mazzarella, B. Brodsky, A. Zagari, and H. M. Berman, “X-ray crystallographic determination of a collagen-like peptide with the repeating sequence (Pro-Pro-Gly),” *J Mol Biol*, vol. 280, pp. 623–638, 1998.
- [29] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, “Charmm: A program for macromolecular energy, minimization, and dynamics calculations,” *J Comput Chem*, vol. 4, pp. 187–217, 1983.
- [30] A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, “All-atom empirical potential for molecular modeling and dynamics studies of proteins,” *J Phys Chem*, vol. 102, pp. 3586–3616, 1998.
- [31] D. Anderson, “Collagen self-assembly: A complementary experimental and theoretical perspective,” Ph.D. dissertation, University of Toronto, 2005.

- [32] A. T. Brünger and M. Karplus, “Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison,” *Proteins*, vol. 4, pp. 148–156, 1988.
- [33] R. Berisio, L. Vitagliano, L. Mazzarella, and A. Zagari, “Crystal structure of the collagen triple helix model [(Pro-Pro-Gly)₁₀]₃,” *Protein Sci*, vol. 11, pp. 262–270, 2002.
- [34] J. K. Rainey and M. C. Goh, “A statistically derived parameterization for the collagen triple-helix,” *Protein Sci*, vol. 11, pp. 2748–2754, 2002.
- [35] H. D. Loof, L. Nilsson, and R. Rigler, “Molecular dynamics simulation of galanin in aqueous and nonaqueous solution,” *J Am Chem Soc*, vol. 114, pp. 4028–4035, 1992.
- [36] J. Engel and H. P. Bachinger, “Structure, stability and folding of the collagen triple helix,” *Top Curr Chem*, vol. 247, pp. 7–33, 2005.
- [37] M. Doi, Y. Nishi, S. Uchiyama, Y. Nishiuchi, T. Nakazawa, T. Ohkubo, and Y. Kobayashi, “Characterization of collagen model peptides containing 4-fluoroproline; (4(S)-fluoroproline-Pro-Gly)₁₀ forms a triple helix, but (4(R)-fluoroproline-Pro-Gly)₁₀ does not,” *J Am Chem Soc*, vol. 125, pp. 9922–9923, 2003.
- [38] P. Raiteri, A. Laio, and M. Parrinello, “Correlations among hydrogen bonds in liquid water,” *Phys Rev Lett*, vol. 93, pp. 087801(1–4), 2004.
- [39] J. Dunitz, “The entropic cost of bound water in crystals and biomolecules,” *Science*, vol. 264, p. 670, 1994.

- [40] F. Song, K. Wisithphrom, J. Zhou, and L. J. Windsor, “Matrix metalloproteinase dependent and independent collagen degradation,” *Front Biosci*, vol. 11, pp. 3100–3120, 2006.
- [41] V. Denisov, K. Venu, J. Peters, H. Horlein, and B. Halle, “Orientational disorder and entropy of water in protein cavities,” *J Phys Chem B*, vol. 101, pp. 9380–9389, 1997.
- [42] S. Fischer and C. S. Verma, “Binding of buried structural water increases the flexibility of proteins,” *Proc Natl Acad Sci USA*, vol. 96, pp. 9613–9615, 1999.
- [43] Y. Harano and M. Kinoshita, “Translational-entropy gain of solvent upon protein folding,” *Biophys J*, vol. 89, pp. 2701–2710, 2005.
- [44] D. Zhang, U. Chippada, and K. Jordan, “Effect of the structural water on the mechanical properties of collagen-like microfibrils: a molecular dynamics study,” *Ann Biomed Eng*, vol. 35, pp. 1216–1230, 2007.
- [45] C. A. Kim and J. M. Berg, “Thermodynamic β -sheet propensities measured using a zinc-finger host peptide,” *Nature*, vol. 362, pp. 267–270, 1993.
- [46] S. Bandyopadhyay, S. Chakraborty, and B. Bagchi, “Coupling between hydration layer dynamics and unfolding kinetics of HP-36,” *J Chem Phys*, vol. 125, pp. 084912(1–11), 2006.
- [47] K. Griebenow and A. M. Klibanov, “On protein denaturation in aqueous-organic mixtures but not in pure organic solvents,” *J Am Chem Soc*, vol. 118, pp. 11 695–11 700, 1996.
- [48] K. Griebenow and A. M. Klibanov, “Lyophilization-induced reversible changes in the secondary structure of proteins,” *Proc Natl Acad Sci USA*, vol. 92, pp.

10 969–10 976, 1995.

- [49] S. Leikin, D. C. Rau, and V. A. Parsegian, “Direct measurement of forces between self-assembled proteins: temperature-dependent exponential forces between collagen triple helices,” *Proc Natl Acad Sci USA*, vol. 91, pp. 276–280, 1994.
- [50] S. Leikin and D. C. Rau and V. A. Parsegian, “Temperature-favoured assembly of collagen is driven by hydrophilic not hydrophobic interactions,” *Nat Struct Biol*, vol. 2, pp. 205–210, 1995.
- [51] S. Leikin, V. A. Parsegian, W. Yang, and G. E. Walrafen, “Raman spectral evidence for hydration forces between collagen triple helices,” *Proc Natl Acad Sci USA*, vol. 94, pp. 11 312–11 317, 1997.
- [52] J. W. Ruberti and N. J. Hallab, “Strain-controlled enzymatic cleavage of collagen in loaded matrix,” *Biochem Biophys Res Commun*, vol. 336, pp. 483–489, 2005.
- [53] C. Huang and I. V. Yannas, “Mechanochemical studies of enzymatic degradation of insoluble collagen fibers,” *J Biomed Mater Res*, vol. 11, pp. 137–154, 1977.
- [54] Y. Nabeshima, E. S. Grood, A. Sakurai, and J. H. Herman, “Uniaxial tension inhibits tendon collagen degradation by collagenase in vitro,” *J Orthop Res*, vol. 14, pp. 123–130, 1996.
- [55] D. Marsolais, E. Duchesne, C. H. Côté, and J. Frenette, “Inflammatory cells do not decrease the ultimate tensile strength of intact tendons in vivo and in vitro: protective role of mechanical loading,” *J Appl Physiol*, vol. 102, pp. 11–17, 2007.

APPENDIX A

CHARACTERIZATION OF UNWINDING

We generated local coordinate triads $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ to characterize the conformation of the triple helix as follows. We constructed triangles joining the C_α atoms of Gly, X, and Y of adjacent chains along the length of the molecule (**Fig. 15 (A)**). The centroid of each triangle was chosen as the local coordinate origin and \hat{e}_1 was defined as the unit vector along the line joining the centroid to the midpoint of the line segment joining C_α atoms of chain-1 and chain-2. Setting \hat{e}_3 perpendicular to the plane fixes $\hat{e}_2 = \hat{e}_3 \times \hat{e}_1$ (**Fig. 15(B)**). The first two triads and the last two triads were omitted in the analysis to reduce end effects (**Fig. 15(C)**). The remaining 24 triads within the middle were grouped into three (I,II,III): domains I (triad 1-4) and III (triad 15-24) are the regular GPO repeats at the N and C termini, while domain II (triad 5-15) is the central imino-poor region.

We characterized orientations of triads using Euler angles. Let $\mathbf{X} = \{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ and $\mathbf{X}' = \{\hat{e}'_1, \hat{e}'_2, \hat{e}'_3\}$ be two adjacent triads (**Fig. 15(B)**). Denoting a Cartesian coordinate basis fixed in space as $\mathbf{E} = \{\hat{x}, \hat{y}, \hat{z}\}$, we have

$$\mathbf{X} = A \mathbf{E}, \quad \mathbf{X}' = B \mathbf{E}, \quad (\text{A.1})$$

where A and B are 3×3 matrices containing the Cartesian components of \mathbf{X} and \mathbf{X}' , respectively. We then have

$$\mathbf{X}' = T \mathbf{X} \quad (T \equiv B A^{-1}). \quad (\text{A.2})$$

On the other hand, \mathbf{X} and \mathbf{X}' are related by Euler transformation matrices by

$$\mathbf{X}' = R_x R_y R_z \mathbf{X} \quad \implies \quad T = R_x R_y R_z, \quad (\text{A.3})$$

where

$$R_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_x & \sin \theta_x \\ 0 & -\sin \theta_x & \cos \theta_x \end{pmatrix},$$

$$R_y = \begin{pmatrix} \cos \theta_y & 0 & \sin \theta_y \\ 0 & 1 & 0 \\ -\sin \theta_y & 0 & \cos \theta_y \end{pmatrix},$$

and

$$R_z = \begin{pmatrix} \cos \theta_z & \sin \theta_z & 0 \\ -\sin \theta_z & \cos \theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

As A and B are easily measurable in a given conformation of the molecule, T can be calculated from Eq. A.2, which in turn gives Euler angles θ_x , θ_y , and θ_z by solving Eq. A.3. If we set \hat{z} along the major axis of the triple helix, changes in θ_z represent the degree of the helical twist (torsion). This can be seen by plotting the torsional map as the difference $\Delta\theta_z$ between a triad under consideration and a reference triad. For example, if we set the first triad as the reference, $\Delta\theta_z$ will monotonically increase with the triad number for a right-handed triple helix, and the triad that reaches $\Delta\theta_z = 360^\circ$ will define the period of the triple helix. Because θ_z is the rotation angle of triads with respect to the z -axis, our method is not affected by the fact that the triangles in **Fig. 15(A)** are not exactly perpendicular to the helical axis. For our analysis, we chose the reference point as the first triad. If there was unwinding of the triple helix in a particular region, it would show up as a change in slope of the torsional map (*cf.* **Fig. 3**).

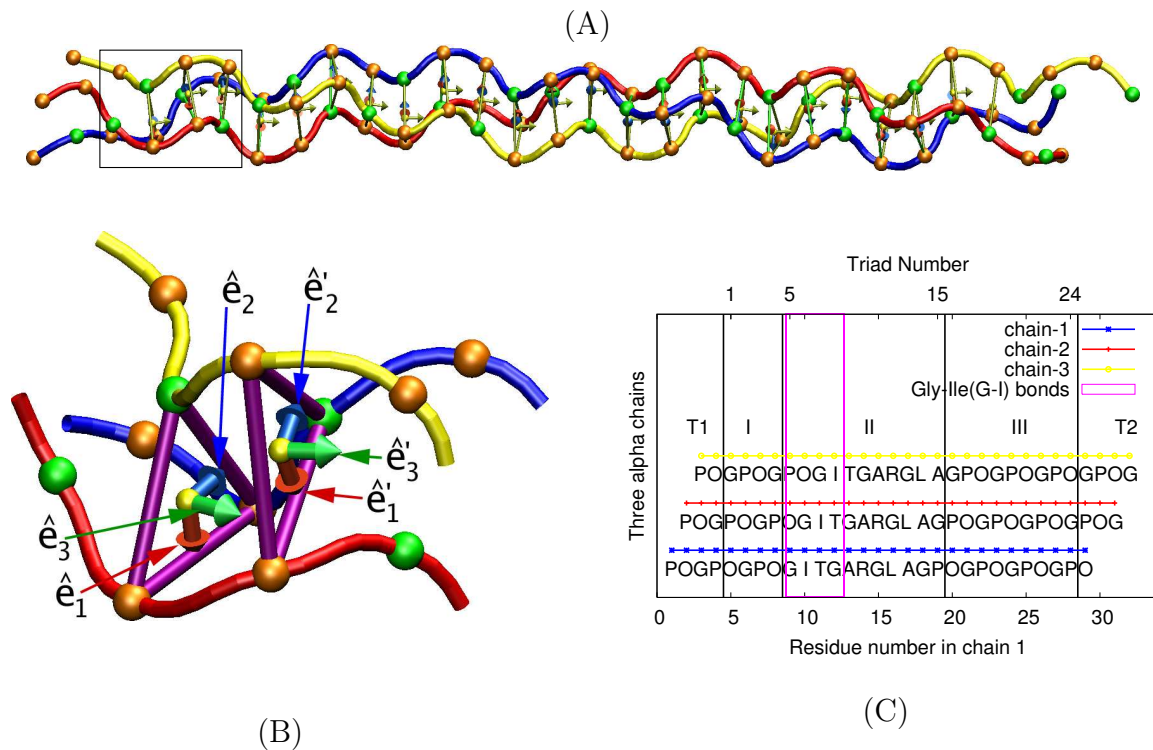


Fig. 15. The three α -chains are shown in blue (chain-1), red (chain-2), and yellow (chain-3). Spheres are C $_{\alpha}$ atoms of Gly (green) and X/Y (orange). (A) Overview of 1BKV with the 24 triads. (B) Magnified view of the box in (A). To show the face of the triangle, the molecule is rotated about the vertical axis of the paper plane. Two neighboring triads are denoted $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ and $\{\hat{e}'_1, \hat{e}'_2, \hat{e}'_3\}$. (C) Triad numbers versus amino acid sequence. The horizontal shift of the three chains reflects their staggered structure in the triple helix. Triads are grouped into domains I, II, and III, as explained in the text. T1 and T2 are unconsidered regions to eliminate end effects. Triads 5 to 8, marked between grey dotted vertical lines, contain Gly-Ile bonds. (Note: The first proline (P) residue in chain-1 in (C) is shown only for clarity and is invisible in the X-ray crystal structure of 1BKV.)

VITA

Name: Krishnakumar Mayuram Ravikumar

Date of Birth: August 13, 1982

Address:

337 Zachry Engineering Center

3120 TAMU

College Station, TX 77843-3120

krishhere@mcbm.tamu.edu

979-204-5014

Education:

M.S. Biomedical Engineering May 2008

Texas A&M University

B.E. Mechanical Engineering May 2004

University of Madras, India

Publications:

Spontaneous Unwinding of Labile Domain in Collagen Triple Helix K. M.

Ravikumar, J. D. Humphrey, and W. Hwang. Journal of Mechanics of Materials
and structures (Vol.2 (6), 2007)

Region-specific role of water in collagen unwinding and assembly K. M. Ravikumar
and W. Hwang. Proteins (in press)